

小説推薦システムの構築に向けた検索表現と書評の分析 Analysis of Query Words and Reviews for the Construction of a Novel Recommendation System

増田 純太[†] 杉本 徹[†]
Junta Masuda Toru Sugimoto

1. はじめに

現在、インターネット上の通販サイトの充実により、パソコンを用いて書籍を購入する人が増えている。しかし通販サイトには膨大な数の商品が存在するため、自分の嗜好に合った書籍を探すのは難しい。また書籍を効率よく探すには、良い検索の方法を熟知していなければならない。

そこで私は、自然言語によるユーザとの対話を通して、小説を推薦することで、ユーザの書籍の選択の手助けとなる「小説推薦システム」の構築を進めている。現在のシステムは、どのような小説を読みたいかを入力すると、システムはジャンルやあらすじに対しキーワード検索を行い、希望に合った小説を提示する。しかし、「明るい小説」や「ワクワクする小説」などの、感性的な表現を含んだ入力には対応していない。

こうした検索表現をシステムが理解するには、まずユーザがどのような検索表現を用いて小説を探すのか調べて、次にそれらの検索表現と各小説の関係を明らかにする必要がある。ユーザが入力する表現に基づいて検索を行うシステムとして、楽曲検索システム MRS-LI がある[1]。この研究では、ユーザが入力にどのような印象表現を用いるか調べるためアンケートを行なっている。またその結果から、システムで使用する印象尺度（形容語で対になっているもの）を決定し、楽曲とこの印象尺度を対応させている。

そこで本研究では、小説に関してもユーザが推薦システム利用時に入力すると予想される検索表現の種類を確認するため、アンケートを実施した。また Web 上にある書評からも、小説に対して利用される検索表現を収集した。本稿では、このアンケート及び書評からの収集の結果を考察し、収集した検索表現を分類する。

2. アンケートによる検索表現の収集

2.1 アンケートの内容

アンケートでは、「あなたが好きな小説にはどのような特徴（印象や雰囲気）がありますか。」というように、その人が小説をどのような観点で捉えているかを問う質問や、「時間が沢山あったら、どのような小説を読みますか。」というように、その人が小説を選ぶ際の基準を問う質問を提示した。

2.2 アンケートの結果

アンケートは 20 代の男性 17 名に行った。始めに、回答者には「1 ヶ月でどのくらい小説を読むか」を質問した。その回答結果を表 1 に示す。

表 1 1 ヶ月に読む小説の数の分布

1 冊未満	4 人	3 冊	4 人
1 冊	2 人	4 冊	1 人
2 冊	3 人	5 冊以上	3 人

この結果から、普段小説をほとんど読まない人から、読書経験が豊富な人まで、様々な人から検索表現を収集できたことが分かった。またこのアンケートから、ユーザが入力すると予想される 154 個の検索表現が収集できた。収集した検索表現の例を表 2 に示す。

表 2 検索表現の例

ほっこり／サスペンス／暗い／推理／ファンタジー／どんより／長編／重たい／リアリティのある／明るい／淡々とした／可愛い／軽い／話題になっている／暖かい／話が長い／ワクワクする

3. 書評からの検索表現の収集

3.1 収集の方法

ショッピングサイト Amazon.co.jp に掲載されている小説の商品レビューから、小説推薦システムの検索表現となりうる表現を収集した。まず、ブックレビューサイト「ブックログ」[2]に登録されている小説から、日本語ウィキペディアの記事「Category:小説のジャンル」で定義されているジャンル名をタグに含む、合計 2,625 冊を抽出し、それらに付与された 59,214 レビューを取得した。

そして、レビュー文にある「小説」、「本」、「作品」、「物語」、「内容」、「ストーリー」という単語を修飾している形容詞及び副詞を検索表現の候補として収集した。

3.2 収集結果

上記の方法で、形容詞が 594 個、副詞が 372 個、合計 966 個収集できた。これにより収集した表現の例を、表 3 と表 4 に示す。

表 3 検索表現の例（形容詞）

面白い／良い／強い／悪い／深い／すごい／切ない／素晴らしい／高い／美しい／長い／うまい／難しい／悲しい／怖い／優しい／暗い／熱い／細かい／恐ろしい／嬉しい／珍しい／激しい

表 4 検索表現の例（副詞）

しっかり／淡々／ワクワク／ドロドロ／がっかり／あっさり／ハラハラ／コテコテ／ドタバタ／ゾクゾク／だらだら／キラキラ／じっくり／バラバラ／ベタベタ／ガックリ／さっぱり／ギスギス

[†] 芝浦工業大学, Shibaura Institute of Technology

4. 検索表現の分析

4.1 アンケートで収集した検索表現の分類と考察

アンケートで収集した 154 個の検索表現を、[1]を参考に以下の基準で分類した。その結果の一部を、表 5 に示す。

表 5 分類結果の一部（アンケート）

タイプ（該当数）	該当例
感性的特徴(30)	暗い／明るい／可愛い／温かい／悲しい／淡々とした
感情反応(8)	どんより／ドキドキ／ほっこり
内容的特徴(40)	ドタバタ／ハッピーエンド／論理的な／躍動感のある
ジャンル(17)	サスペンス／推理
感想・評論(5)	読み応えがある／つまらない
認知度(1)	話題になっている
構成(7)	短編／長編／話が長い
その他(46)	洗練された文章／メディアミックス／主人公が強い

「感性的特徴」には「可愛い」や「淡々とした」などの、小説が持っている雰囲気を説明した表現を分類した。ここに分類された表現の中には、接頭に「かなり」や「少し」などの、程度を表す単語と一緒に用いられることが多かった。推薦システムがこのような入力に対応するためには、程度の量を数値化し[3]、小説に対応づける必要がある。

「感情反応」には「どんより」や「ほっこり」などの、読者の感情状態の変化を表す言葉を分類した。なお、「感性的特徴」に分類した「明るい」や「暗い」といった言葉は「感情反応」に分類されそうだが、明るい小説を読んでも読者の気分が明るくなるとは限らないため、これらの表現は「感性的特徴」に分類した。

「内容的特徴」には「ドタバタ」や「ハッピーエンド」などの、小説の内容に関して説明している表現を分類した。「その他」を除き、最も多くの検索表現がこのタイプに分類された。現在のシステムでは、このような表現に対してはあらずじに対するキーワード検索で対応しているが、同じ意味だがあらずじにない表現や、「論理的な」など小説を読まないとは分からないような表現に対応させる必要がある。

「ジャンル」には「サスペンス」や「推理」などの表現を分類した。これらの表現に対する検索機能は、現在のシステムで実装済みである。また「感想・評論」には「読み応えがある」や「つまらない」などの、読者の小説に対する感想や評論を分類した。ある小説がどの程度面白いかは読者によって変わるため、推薦システムがこれらの入力に対応するためにはユーザの嗜好をモデル化する必要がある。

さらに「認知度」には知名度を示す表現を分類し、「構成」には小説のページ数や物語の展開を表す表現を分類した。また「その他」には上記に当てはまらないものを分類した。この中には例えば文章の表現方法、コンテンツの販売方法、登場人物に関する表現などが分類されている。

4.2 書評から収集した検索表現の分類と考察

書評から収集した 966 個の検索表現を、アンケートと同じ基準で分類した。その結果の一部を、表 6 に示す。

表 6 分類結果の一部（書評）

タイプ（該当数）	該当例
感性的特徴(179)	美しい／明るい／淡々／おどろおどろ／キラキラ
感情反応(60)	嬉しい／明るく／寂しく／ハラハラ／ゾクゾク
内容的特徴(122)	甘酸っぱい／熱い／あっさり／コテコテ／ドタバタ
感想・評論(107)	面白い／良い／悪い／凄い／がっかり／うまい
構成(35)	長い／あつけない／だらだら／じっくり
その他(461)	高い／安い／比較的／少なくとも／まさに／これぞ

検索表現の収集条件を、3.1 節にあげた 6 単語を修飾している形容詞及び副詞としたため、「ジャンル」や「認知度」に分類される表現はなかった。

アンケート結果と比べて「感情反応」には、「明るく」や「寂しく」など、「感性的特徴＋〜く」のように、副詞化した表現が多く見られた。

また「その他」には、値段に関しての表現や、他の作品と比べた表現が見られた。しかしここに分類された副詞は、2 語以上で用いられる表現などの誤抽出が多いので、同時に用いられる名詞や動詞も考慮するなど、収集条件を見直す必要がある。

5. まとめ

本稿では、小説推薦システム利用時にユーザが入力すると予想される検索表現の種類を、アンケート調査と書評の分析を通して明らかにした。これにより、小説の内容に関する表現だけでなく、感性的な特徴や読者の反応に関する表現など、様々な検索表現が使われる可能性があることが分かった。

今後は、今回の実験で得られた表現と小説を対応させる方法を考えていく。熊本らの研究では、「感性的特徴」と「感情反応」に当たるものを利用して楽曲の印象を決定しているが、小説の場合は「内容的特徴」も重要であるため、これも加味した、小説の印象の定式化を行っていきたい。そして最終的には、感性的な表現を含んだ入力に対応する小説推薦システムを構築する。

謝辞

本研究を進めるにあたり、アンケートに快くご協力いただいた皆様に感謝いたします。

参考文献

- [1] 熊本忠彦, 大田公子, “印象に基づく楽曲検索：検索表現の収集と分析”, 情処研報(自然言語処理), Vol.2001, No.112, pp.101-106
- [2] 「ブックログ」, <<http://booklog.jp/>>, (2011/7/27 時点)
- [3] 織田 揮準, “日本語の程度量表現用語に関する研究”, 教育心理学研究 18(3), pp.166-176 (1970)